



Available online at : <http://bit.ly/InfoTekJar>

InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan

ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600



Data Mining

Analisis Pola Penyakit Hipertensi Menggunakan Algoritma C4.5

Nurul Azwanti, Erlin Elisa

Universitas Putera Batam, Jalan R. Soeprapto-Muka Kuning, Batam

KEYWORDS

Data mining, algoritma C4.5, hypertension

CORRESPONDENCE

Phone: +62 85270085861

E-mail: nurulazw@rocketmail.com

ABSTRACT

There are approximately 95% of cases of unknown cause of hypertension, while the rest caused by other diseases such as coronary heart disease, impaired kidney function, and impaired cognitive function or stroke. RSUD Embung Fatimah is an Indonesian hospital located in Batam Island Riau Province. In 2015, the total number of inpatients for hospitalization reaches 10,317 inhabitants. With the large number of patients per year it causes patient data is increasing. To overcome the problem in tackling people with hypertension disease, it is necessary to analyze the existing disease data, to predict the patient's illness which must be handled based on the pattern of the disease. In data mining there is a model that can be used to predict a pattern in a condition that is predictive or prediction model. One of the algorithms that can be used to create a decision tree (decision tree) is the C4.5 algorithm. The C4.5 algorithm is a method used for predictive classification. Using C4.5 algorithm method, the researcher can classify the pattern of hypertension as a comorbid illness of heart failure, kidney failure, diabetes, stroke and hypoglycemia. In this study, researchers used WEKA (Waikato Environment for Knowledge Analysis) software as tools or tools used to perform testing in order to obtain the pattern of disease from hypertension. From the research findings in the find that in the prediction of hypertension disease as a disease, the attributes that are very influential to hypertension are heart failure.

PENDAHULUAN

Hipertensi merupakan salah satu jenis penyakit yang menyerang organ dalam tubuh yang dapat membahayakan hidup seseorang. Di Negara maju prevalensi mencapai 37,3% [1]. Terdapat sekitar 95% kasus hipertensi yang tidak diketahui penyebabnya, sedangkan sisanya ditimbulkan akibat adanya penyakit lain seperti penyakit jantung koroner, gangguan fungsi ginjal, dan gangguan fungsi kognitif atau stroke [2]. Di Indonesia masalah hipertensi cenderung meningkat (Rahajeng & Tuminah, 2009: 581). Prevalensi penyakit hipertensi di Indonesia rata-rata pada penduduk usia 25-70 tahun sebesar 25,8%, namun cakupan kasus oleh tenaga kesehatan hanya sebesar 36,8% dari total penderita yang diperkirakan dan sisanya tidak terdiagnosis dengan baik di masyarakat [3].

RSUD Embung Fatimah adalah rumah sakit Indonesia yang terletak di Provinsi Kepulauan Riau Kota Batam. Di tahun 2015, jumlah total pasien untuk rawat inap mencapai 10.317 jiwa (Profil kesehatan Kota Batam, 2015: 231). Dengan komplikasi beberapa penyakit yang muncul secara bersama, Dengan jumlah pasien penderita hipertensi yang cukup banyak, tidak menutup kemungkinan dapat mengakibatkan bertambahnya angka kematian akibat hipertensi. Hal tersebut dikarenakan hipertensi dapat menjadi faktor penyerta penyakit – penyakit lainnya seperti, gagal ginjal, gagal jantung, diabetes, stroke dan

hipoglikemia. Dengan banyaknya jumlah pasien pertahun nya menyebabkan data pasien semakin bertambah. Saat ini data – data tersebut disimpan pada aplikasi SIMRS (Sistem Informasi Manajemen Rumah Sakit) dan belum dimanfaatkan secara maksimal. Kurangnya perhatian terhadap tumpukan data pasien ini dapat menyebabkan banyaknya kendala atau permasalahan yang seharusnya dapat di antisipasi jauh sebelum permasalahan tersebut muncul. Semakin lama data pasien yang disimpan tersebut dibiarkan tentunya akan semakin sulit memprediksikan sebuah solusi untuk memecahkan permasalahan apabila permasalahan tersebut muncul. Selain itu, data hanya akan menjadi sebuah tumpukan arsip atau file yang disimpan sebagai formalitas kerja manajemen rumah sakit saja. Oleh sebab itu peneliti melakukan penelitian dengan memanfaatkan tumpukan data tersebut agar dapat dimanfaatkan secara maksimal.

Untuk mengatasi masalah dalam menanggulangi penderita penyakit Hipertensi, diperlukan analisa terhadap data penyakit yang ada, untuk diprediksi penyakit pasien yang harus ditangani berdasarkan pola penyakit. Data mining digunakan untuk mengetahui pola data, mengatur hubungan informasi tersembunyi, struktur aturan Asosiasi, memperkirakan item tidak diketahui, nilai-nilai untuk mengklasifikasikan benda, menyusun cluster objek homogen, dan mengungkap berbagai temuan yang tidak mudah diproduksi oleh CBIS (*computer based information system*) klasik. Dengan demikian, hasil DM mewakili dukungan yang berharga untuk membuat keputusan

[4]. Dalam data mining ada sebuah model yang dapat digunakan untuk memprediksikan sebuah pola dalam suatu kondisi yaitu model prediktif atau prediksi. Model prediktif sering menerapkan pembelajaran diawasi fungsi untuk memperkirakan nilai-nilai tidak diketahui atau masa depan tergantung variabels berdasarkan fitur terkait variabel independen. [4]. Untuk memprediksikan suatu kondisi dalam data mining, diperlukan sebuah alat bantu yang mampu mengolah data-data menjadi informasi yang akan digunakan untuk memprediksikan dan memuatnya menjadi sebuah pola. Pohon keputusan memperlihatkan faktor-faktor kemungkinan atau probabilitas yang akan mempengaruhi alternatif-alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil alternatif keputusan tersebut [5]. Salah satu algoritma yang dapat digunakan untuk membuat pohon keputusan (*decission tree*) adalah algoritma C4.5.

LANDASAN TEORI

Hipertensi

Hipertensi merupakan salah satu penyakit degeneratif yang perlu diwaspadai karena merupakan penyebab kematian utama ketiga di Indonesia untuk semua umur (6,8%), setelah stroke (15,4%) dan tuberculosis (7,5%) (Anisa, Bangun, & Sinulingga, 2014:). Pada penelitian [7] disebutkan, bahwa hipertensi merupakan gangguan sistem peredaran darah yang ditandai oleh meningkatnya tekanan darah dalam tubuh. Seseorang yang menderita hipertensi dapat berpotensi mengalami komplikasi, seperti stroke dan penyakit jantung lainnya yang dapat berakibat fatal. Kemudian dalam penelitian [8] menurut WHO, Hipertensi atau "tekanan darah tinggi" adalah tekanan darah sistole ≥ 140 mmHg dan tekanan darah diastole ≥ 90 mmHg. Tekanan darah sistole adalah suatu kondisi penguncupan jantung akibat kontraksi otot jantung, sedangkan tekanan darah diastole adalah sebuah kondisi pengenduran otot jantung pada waktu pengisian kembali jantung oleh darah [9].

Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui (Pramadhani & Setiadi, 2014: 832). Sedangkan dalam Penelitian (Yuli, 2017: 214), menurut Gartner Group, data mining adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistik dan matematika.

Alur kerja dari Data Mining berisi beberapa tahap, yaitu :

1. *Data Collection*

Pada tahap pengumpulan, data sering disimpan di database atau data *warehouse* untuk selanjutnya di proses.

2. *Feature extraction and data cleaning*

Ketika data sudah terkumpul, banyak bentuk data yang tidak cocok digunakan untuk diolah. Untuk membuat data sesuai dengan format agar data dapat diolah, penting untuk mengubah data menjadi format yang sesuai dengan algoritma data mining. Hasil akhir dari prosedur ini adalah sebuah kumpulan data

set yang terstruktur, yang dapat secara efektif digunakan oleh program komputer.

3. *Analytical processing and algorithms*

Bagian akhir dari proses Data Mining adalah untuk merancang metode analisis yang efektif dari data yang diolah.

4. *Preprocessing*

Preprocessing adalah suatu proses atau langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas serta menjadi input yang baik bagi data mining tools. Tahap *preprocessing* terbagi ke dalam beberapa tahap, yaitu :

- a. *Data Cleaning*

Data Cleaning berfungsi untuk membersihkan data dengan mengisi nilai – nilai yang hilang, *smoothing noise* data, mengidentifikasi dan menghapus *outlier*, serta menyelesaikan inkonsisten

- b. *Data Integration*

Data Integration merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. *Data integration* dapat membantu mengurangi dan menghindari redundansi data dan inkonsistensi dalam kumpulan data yang dihasilkan. Ini dapat membantu meningkatkan akurasi dan kecepatan proses data mining berikutnya.

- c. *Data Reduction*

Data Reduction dapat diterapkan untuk mengurangi ukuran data dengan menggabungkan atau menghilangkan data yang tidak dibutuhkan.

- d. *Data Transformation*

Data Transformation berfungsi mengubah data ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan [11]

Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang sangat populer yang digunakan oleh banyak peneliti di dunia, hal ini dijelaskan oleh Xindong Wu dan Vipin Kumar dalam bukunya yang berjudul *The Top Ten Algorithms in Data Mining*. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang di ciptakan oleh J. Rose Quinlan (Yuli, 2017: 216). Algoritma C4.5 mempunyai kelebihan karena dapat menghasilkan model berupa pohon (Andri, Kunang, & Murniati, 2013: 58).

Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

- a. Pilih atribut sebagai akar
- b. Buat cabang untuk masing-masing nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama [13].

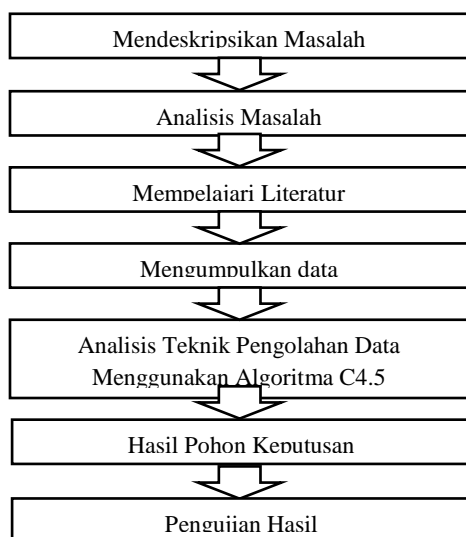
Algoritma C4.5 digunakan untuk membangun sebuah pohon keputusan yang mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar pohon keputusan. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan [14].

Waikato Environment for Knowledge Analysis (WEKA)

WEKA adalah sebuah paket tools machine learning praktis. WEKA merupakan singkatan dari *Waikato Environment for Knowledge Analysis*, yang dibuat di Universitas Waikato, New Zealand untuk penelitian, pendidikan dan berbagai aplikasi. WEKA mampu menyelesaikan masalah-masalah data mining di dunia nyata, khususnya klasifikasi yang mendasari pendekatan-pendekatan machine learning. Perangkat lunak ini ditulis dalam hirarki class Java dengan metode berorientasi objek dan dapat berjalan hampir di semua platform. WEKA mudah digunakan dan diterapkan pada beberapa tingkatan yang berbeda. Tersedia implementasi algoritma-algoritma pembelajaran *state-of-the-art* yang dapat diterapkan pada dataset dari command line. WEKA mengandung tools untuk *pre-processing* data, klasifikasi, regresi, clustering, aturan asosiasi, dan visualisasi. User dapat melakukan *preprocess* pada data, memasukkannya dalam sebuah skema pembelajaran, dan menganalisa classifier yang dihasilkan dan performansinya semua itu tanpa menulis kode program sama sekali. Ada banyak metodologi Data Mining yang dapat dilakukan pada WEKA, salah satu yang populer adalah pohon keputusan (*decision tree*). Pohon keputusan merupakan salah satu metode klasifikasi yang sangat menarik yang melibatkan konstruksi pohon keputusan yang terdiri dari node keputusan yang dihubungkan dengan cabang-cabang dari simpul akar sampai ke node daun (akhir). Pada node keputusan atribut akan diuji, dan setiap hasil akan menghasilkan cabang. Setiap cabang akan diarahkan ke node lain atau ke node akhir untuk menghasilkan suatu keputusan [15].

METODE

Adapun metode penelitian digambarkan pada gambar 1 berikut ini



Gambar 1. Metode Penelitian

HASIL PENELITIAN

Dalam penelitian ini peneliti meneliti pola penyakit pasien dengan memanfaatkan data mining untuk pola atau informasi menarik pada data terpilih yang dikelompokkan dalam klasifikasi variabel yang bersifat kategorik untuk mengklasifikasikan

penyakit pasien [16] dalam dua keputusan, yaitu hipertensi ringan dan hipertensi berat berdasarkan usia, jenis kelamin, lama dirawat, diabetes, hipoglikemia, gagal jantung, stroke, gagal ginjal.

Pemilihan variabel

Langkah pertama adalah langkah terpenting dalam penelitian yaitu penentuan variabel yang akan diteliti dengan mendefinisikan variabel yang akan di olah menjadi sebuah pohon keputusan. Peneliti mencoba dan mencari cara untuk menggambarkan pola dan kecenderungan dalam data sehingga mendapatkan pohon keputusan untuk mengklasifikasi pola penyakit Hipertensi kedalam kategori hipertensi ringan atau hipertensi berat berdasarkan variabel yang akan diolah. Berikut pengklasifikasian data berdasarkan variabel dalam penelitian adalah sebagai berikut :

1. Usia dikategorikan berdasarkan range dinas kesehatan menurut Depkes RI (2009) dapat diklasifikasikan sebagai berikut:

Tabel 1. Kategori Usia

Umur	Klasifikasi
0 – 5 Tahun	Balita
6 – 11 Tahun	Kanak – kanak
12 – 16 Tahun	Remaja Awal
17 – 25 Tahun	Remaja Akhir
26 – 35 Tahun	Dewasa Awal
36 – 45 Tahun	Dewasa Akhir
46 – 55 Tahun	Lansia Awal
56 – 65 Tahun	Lansia Akhir
>65 Tahun	Manula

2. Jenis Kelamin dapat diklasifikasi menjadi Pria dan Wanita.
3. Lama dirawat dirumah sakit di simbolkan dengan A, B, C, D, E, F dan Pengelompokan lama rawatan di klasifikasikan dengan rumus Sturgess:

$$\begin{aligned} \text{Jumlah Kelas} &= 1 + 3.3 \log n \text{ (jumlah data)} \\ &= 1 + 3.3 \log (34) \\ &= 1 + 3.3 (1.53147) \\ &= 6.0538 \text{ atau } 6 \end{aligned}$$

Tabel 1. Tabel Lama Rawatan

Umur	Klasifikasi
1 – 6 Hari	A
7 – 12 Hari	B
13 – 18 Hari	C
19 - 24 Hari	D
25 – 30 Hari	E
31 – 36 > Hari	F

4. Diabetes diklasifikasikan dengan apakah pasien mengidap penyakit diabetes, dan dibatasi dengan Ya atau Tidak

5. hipoglikemia diklasifikasikan dengan apakah pasien mengidap penyakit hipoglikemia, dan dibatasi dengan Ya atau Tidak
6. Gagal Jantung diklasifikasikan dengan apakah pasien mengidap penyakit gagal jantung, dan dibatasi dengan Ya atau Tidak
7. Stroke diklasifikasikan dengan apakah pasien mengidap penyakit stroke, dan dibatasi dengan Ya atau Tidak
8. Gagal ginjal diklasifikasikan dengan apakah pasien mengidap penyakit gagal ginjal, dan dibatasi dengan Ya atau Tidak.
9. Pada Keputusan diklasifikasikan menjadi Hipertensi berat atau hipertensi ringan dan dibatasi dengan ya atau tidak.

Langkah dalam pembentukan pohon keputusan menggunakan Algoritma C4.5 untuk menyelesaikan masalah pada penelitian ini. Yaitu dengan merumuskan :

1. Menghitung jumlah kasus untuk keputusan Hipertensi Ringan dan jumlah keputusan untuk keputusan Hipertensi Berat.
2. Menghitung Entropy dari semua kasus berdasarkan atribut : Usia, jenis kelamin, lama dirawat dan klasifikasi penyakit Diabetes, hipoglikemia, gagal jantung, stroke, gagal ginjal, hipovolemia, dengan rumus sebagai berikut :

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i \quad (1)$$

Setelah itu dilakukan perhitungan Gain untuk setiap atribut.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Melakukan Pra – proses

Berdasarkan variabel – variabel yang sudah terpilih, format akan menjadi seperti yang terlihat pada gambar dibawah ini.

No	Keterangan Usia	Jenis Kelamin	Lama Dirawat	Diabetes	Gagal Ginjal	Gagal Jantung	Stroke	Hipoglikemia	Keputusan
1	Lansia Awal	Pria	A	Ya	Tidak	Tidak	Tidak	Tidak	RINGAN
2	Lansia Akhir	Wanita	A	Ya	Tidak	Ya	Ya	Tidak	BERAT
3	Lansia Awal	Pria	B	Tidak	Tidak	Ya	Ya	Ya	BERAT
4	Remaja Akhir	Wanita	A	Tidak	Tidak	Ya	Tidak	Ya	RINGAN
5	Dewasa Akhir	Pria	B	Tidak	Tidak	Ya	Ya	Ya	BERAT
6	Lansia Akhir	Pria	A	Tidak	Ya	Tidak	Tidak	Ya	RINGAN
7	Lansia Awal	Wanita	A	Tidak	Ya	Tidak	Tidak	Ya	RINGAN
8	Dewasa Akhir	Wanita	B	Tidak	Ya	Tidak	Tidak	Ya	BERAT
9	Lansia Awal	Pria	A	Ya	Tidak	Tidak	Tidak	Tidak	RINGAN
10	Mamula	Wanita	A	Ya	Ya	Tidak	Ya	Tidak	RINGAN
11	Lansia Akhir	Wanita	A	Ya	Tidak	Ya	Ya	Tidak	BERAT
12	Lansia Awal	Pria	C	Ya	Tidak	Ya	Ya	Tidak	BERAT
13	Dewasa Akhir	Pria	B	Ya	Tidak	Ya	Tidak	Tidak	RINGAN
14	Lansia Awal	Pria	C	Ya	Ya	Ya	Tidak	Tidak	BERAT
15	Mamula	Pria	A	Tidak	Tidak	Tidak	Tidak	Ya	RINGAN
16	Lansia Awal	Wanita	B	Tidak	Tidak	Ya	Ya	Tidak	RINGAN
17	Remaja Akhir	Pria	A	Tidak	Tidak	Ya	Ya	Ya	BERAT

Gambar 2. Keputusan Hipertensi

Menghitung Atribut Akar

Berikut ini adalah penjelasan lebih terperinci mengenai tiap – tiap langkah dalam pembentukan pohon keputusan.

Menghitung Nilai Entropy tiap – tiap atribut

$$Entropy(Total) = \left(-\frac{9}{17} * \log_2 \left(\frac{9}{17} \right) \right) + \left(-\frac{8}{17} * \log_2 \left(\frac{8}{17} \right) \right) = 0,9975$$

Atribut Usia

- Entropy(Remaja Akhir) = $\left(-\frac{1}{2} * \log_2 \left(\frac{1}{2} \right) \right) + \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2} \right) \right) = 1$
- Entropy(Dewasa awal) = $\left(-\frac{0}{0} * \log_2 \left(\frac{0}{0} \right) \right) + \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0} \right) \right) = 0,0$
- Entropy(Dewasa Akhir) = $\left(-\frac{1}{3} * \log_2 \left(\frac{1}{3} \right) \right) + \left(-\frac{2}{3} * \log_2 \left(\frac{2}{3} \right) \right) = 0,9183$
- Entropy(Lansia Awal) = $\left(-\frac{4}{7} * \log_2 \left(\frac{4}{7} \right) \right) + \left(-\frac{3}{7} * \log_2 \left(\frac{3}{7} \right) \right) = 0,9852$

$$\begin{aligned} \text{e. Entropy(Lansia Akhir)} &= \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) = 0,9183 \\ \text{f. Entropy(Manula)} &= \left(-\frac{2}{2} * \log_2\left(\frac{2}{2}\right)\right) + \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) = 0,0 \end{aligned}$$

Atribut Lama rawatan

$$\begin{aligned} \text{a. Entropy(A)} &= \left(-\frac{7}{10} * \log_2\left(\frac{7}{10}\right)\right) + \left(-\frac{3}{10} * \log_2\left(\frac{3}{10}\right)\right) = 0,8813 \\ \text{b. Entropy(B)} &= \left(-\frac{2}{5} * \log_2\left(\frac{2}{5}\right)\right) + \left(-\frac{3}{5} * \log_2\left(\frac{3}{5}\right)\right) = 0,9710 \\ \text{c. Entropy(C)} &= \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} * \log_2\left(\frac{2}{2}\right)\right) = 0,0 \\ \text{d. Entropy(D)} &= \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0,0 \\ \text{e. Entropy(E)} &= \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0,0 \\ \text{f. Entropy(F)} &= \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0,0 \end{aligned}$$

Atribut Jenis kelamin

$$\text{Entropy(Pria)} = \left(-\frac{5}{10} * \log_2\left(\frac{5}{10}\right)\right) + \left(-\frac{5}{10} * \log_2\left(\frac{5}{10}\right)\right) = 1 = 0,9852$$

Atribut Diabetes

$$\begin{aligned} \text{a. Entropy(Tidak)} &= \left(-\frac{5}{9} * \log_2\left(\frac{5}{9}\right)\right) + \left(-\frac{4}{9} * \log_2\left(\frac{4}{9}\right)\right) = 0,9911 \\ \text{b. Entropy(Ya)} &= \left(-\frac{4}{8} * \log_2\left(\frac{4}{8}\right)\right) + \left(-\frac{4}{8} * \log_2\left(\frac{4}{8}\right)\right) = 1 \end{aligned}$$

Atribut Gagal ginjal

$$\begin{aligned} \text{a. Entropy(Tidak)} &= \left(-\frac{6}{12} * \log_2\left(\frac{6}{12}\right)\right) + \left(-\frac{6}{12} * \log_2\left(\frac{6}{12}\right)\right) = 1 \\ \text{b. Entropy(Ya)} &= \left(-\frac{3}{5} * \log_2\left(\frac{3}{5}\right)\right) + \left(-\frac{2}{5} * \log_2\left(\frac{2}{5}\right)\right) = 0,9710= \end{aligned}$$

Atribut Gagal jantung

$$\begin{aligned} \text{a. Entropy(Tidak)} &= \left(-\frac{6}{7} * \log_2\left(\frac{6}{7}\right)\right) + \left(-\frac{1}{7} * \log_2\left(\frac{1}{7}\right)\right) = 0,5917 \\ \text{b. Entropy(Ya)} &= \left(-\frac{3}{10} * \log_2\left(\frac{3}{10}\right)\right) + \left(-\frac{7}{10} * \log_2\left(\frac{7}{10}\right)\right) = 0,8813 \end{aligned}$$

Atribut Stroke

$$\begin{aligned} \text{a. Entropy(Tidak)} &= \left(-\frac{7}{9} * \log_2\left(\frac{7}{9}\right)\right) + \left(-\frac{2}{9} * \log_2\left(\frac{2}{9}\right)\right) = 0,7642 \\ \text{b. Entropy(Ya)} &= \left(-\frac{2}{8} * \log_2\left(\frac{2}{8}\right)\right) + \left(-\frac{6}{8} * \log_2\left(\frac{6}{8}\right)\right) = 0,8113 \end{aligned}$$

Atribut Hipoglikemia

$$\begin{aligned} \text{a. Entropy(Tidak)} &= \left(-\frac{5}{9} * \log_2\left(\frac{5}{9}\right)\right) + \left(-\frac{4}{9} * \log_2\left(\frac{4}{9}\right)\right) = 0,9911 \\ \text{Entropy(Ya)} &= \left(-\frac{4}{8} * \log_2\left(\frac{4}{8}\right)\right) + \left(-\frac{4}{8} * \log_2\left(\frac{4}{8}\right)\right) = 1 \end{aligned}$$

Menghitung Nilai Gain tiap – tiap atribut

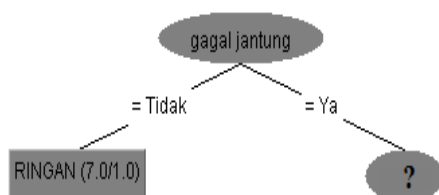
$$\begin{aligned} \text{a. Gain (Total, Usia)} &= 0,9975 - \left(\left(\frac{2}{17} * 1\right) + \left(\frac{0}{17} * 0,0\right) + \left(\frac{3}{17} * 0,9183\right) + \left(\frac{7}{17} * 0,9852\right) + \left(\frac{3}{17} * 0,9183\right) + \left(\frac{2}{17} * 0,0\right)\right) = 0,1501 \\ \text{b. Gain (Total, Lama rawatan)} &= 0,9975 - \left(\left(\frac{10}{17} * 0,8813\right) + \left(\frac{5}{17} * 0,9710\right) + \left(\frac{2}{17} * 0,0\right) + \left(\frac{0}{17} * 0,0\right) + \left(\frac{0}{17} * 0,0\right) + \left(\frac{0}{17} * 0,0\right)\right) = 0,1935 \\ \text{c. Gain (Total, Jenis kelamin)} &= 0,9975 - \left(\left(\frac{10}{17} * 1\right) + \left(\frac{7}{17} * 0,9852\right)\right) = 0,0036 \\ \text{d. Gain (Total, Diabetes)} &= 0,9975 - \left(\left(\frac{9}{17} * 0,9911\right) + \left(\frac{8}{17} * 1\right)\right) = 0,0022 \\ \text{e. Gain (Total, Gagal ginjal)} &= 0,9975 - \left(\left(\frac{12}{17} * 1\right) + \left(\frac{5}{17} * 0,9710\right)\right) = 0,0060 \\ \text{f. Gain (Total, Gagal Jantung)} &= 0,9975 - \left(\left(\frac{7}{17} * 0,5917\right) + \left(\frac{10}{17} * 0,8813\right)\right) = 0,2355 \\ \text{g. Gain (Total, Stroke)} &= 0,9975 - \left(\left(\frac{9}{17} * 0,7642\right) + \left(\frac{8}{17} * 0,8113\right)\right) = 0,2111 \\ \text{h. Gain (Total, Hipoglikemia)} &= 0,9975 - \left(\left(\frac{9}{17} * 0,9911\right) + \left(\frac{8}{17} * 1\right)\right) = 0,0022 \end{aligned}$$

Dari hasil perhitungan diatas dinamakan *Node 1*, nilai Entropy masing – masing atribut yang telah dilakukan menghasilkan nilai Gain, seperti yang terlihat pada tabel

Node	Atribut	Jumlah Kasus(S)	RINGAN	BERAT	Entropy	Gain
1	TOTAL	17	9	8	0.9975	
	Usia					
	Remaja Akhir	2	1	1	1.0000	0.1501
	Dewasa awal	0	0	0	0.0000	
	Dewasa Akhir	3	1	2	0.9183	
	Lansia Awal	7	4	3	0.9852	
	Lansia Akhir	3	1	2	0.9183	
	Manula	2	2	0	0.0000	
	Lama rawatan					
	A	10	7	3	0.8813	0.1935
	B	5	2	3	0.9710	
	C	2	0	2	0.0000	
	D	0	0	0	0.0000	
	E	0	0	0	0.0000	
	F	0	0	0	0.0000	
	Jenis kelamin					
	Pria	10	5	5	1.0000	0.0036
	Wanita	7	4	3	0.9852	
	Diabetes					
	Tidak	9	5	4	0.9911	0.0022
	Ya	8	4	4	1.0000	
	Gagal ginjal					
	Tidak	12	6	6	1.0000	0.0060
	Ya	5	3	2	0.9710	
	Gagal jantung					
	Tidak	7	6	1	0.5917	0.2355
	Ya	10	3	7	0.8813	
	Stroke					
	Tidak	9	7	2	0.7642	0.2111
	Ya	8	2	6	0.8113	
	Hipoglikemia					
	Tidak	9	5	4	0.9911	0.0022
	Ya	8	4	4	1.0000	

Gambar 3. Perhitungan Node 1

Pada perhitungan Node 1 diatas diketahui Gain tertinggi adalah Gagal jantung dengan nilai tertinggi yaitu 0,2880. Dari perhitungan diatas dapat digambarkan sebagai gambar 3.



Gambar 4. Pohon Keputusan Node 1

Menghitung Tree dengan perhitungan Node 1.1

Setelah dilakukan perhitungan node 1, diketahui atribut gagal jantung sebagai atribut tertinggi yaitu sebesar 0,2880. Dengan demikian Gagal jantung dapat menjadi node akar.

Menghitung Nilai Entropy tiap – tiap atribut

$$\text{Entropy (Total)} = \left(-\frac{3}{10} * \log_2 \left(\frac{3}{10}\right)\right) + \left(-\frac{7}{10} * \log_2 \left(\frac{7}{10}\right)\right) = 0,8813$$

Atribut Usia

$$\text{b. Entropy(Remaja Akhir)} = \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2}\right)\right) = 1$$

$$\text{c. Entropy(Dewasa awal)} = \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) = 0,0$$

$$\text{d. Entropy(Dewasa Akhir)} = \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2}\right)\right) = 1$$

$$\text{e. Entropy(Lansia Awal)} = \left(-\frac{1}{4} * \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} * \log_2 \left(\frac{3}{4}\right)\right) = 0,8113$$

$$\text{f. Entropy(Lansia Akhir)} = \left(-\frac{0}{2} * \log_2 \left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} * \log_2 \left(\frac{2}{2}\right)\right) = 0,0$$

$$\text{g. Entropy(Manula)} = \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) = 0,00$$

Atribut Lama rawatan

$$\text{a. Entropy(A)} = \left(-\frac{1}{4} * \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} * \log_2 \left(\frac{3}{4}\right)\right) = 0,8113$$

$$\text{b. Entropy(B)} = \left(-\frac{2}{4} * \log_2 \left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} * \log_2 \left(\frac{2}{4}\right)\right) = 1$$

$$\text{c. Entropy(C)} = \left(-\frac{0}{2} * \log_2 \left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} * \log_2 \left(\frac{2}{2}\right)\right) = 0,0$$

$$\text{d. Entropy(D)} = \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) = 0,0$$

$$\text{e. Entropy(E)} = \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) = 0,0$$

$$\text{f. Entropy(F)} = \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) = 0,0$$

Atribut Jenis kelamin

$$\text{a. Entropy(Pria)} = \left(-\frac{1}{6} * \log_2 \left(\frac{1}{6}\right)\right) + \left(-\frac{5}{6} * \log_2 \left(\frac{5}{6}\right)\right) = 0,6500$$

$$\text{b. Entropy(Wanita)} = \left(-\frac{2}{4} * \log_2 \left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} * \log_2 \left(\frac{2}{4}\right)\right) = 1$$

Atribut Diabetes

$$\text{a. Entropy(Tidak)} = \left(-\frac{2}{5} * \log_2 \left(\frac{2}{5}\right)\right) + \left(-\frac{3}{5} * \log_2 \left(\frac{3}{5}\right)\right) = 0,9710$$

$$b. \text{Entropy}(Y_a) = \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) = 0,7219$$

Atribut Gagal ginjal

$$a. \text{Entropy}(\text{Tidak}) = \left(-\frac{3}{9} * \log_2\left(\frac{3}{9}\right)\right) + \left(-\frac{6}{9} * \log_2\left(\frac{6}{9}\right)\right) = 0,9183$$

$$b. \text{Entropy}(Y_a) = \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) = 0,0$$

Atribut Stroke

$$a. \text{Entropy}(\text{Tidak}) = \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) + \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) = 0,9183$$

$$b. \text{Entropy}(Y_a) = \left(-\frac{1}{7} * \log_2\left(\frac{1}{7}\right)\right) + \left(-\frac{6}{7} * \log_2\left(\frac{6}{7}\right)\right) = 0,5917$$

Atribut Hipoglikemia

$$a. \text{Entropy}(\text{Tidak}) = \left(-\frac{2}{6} * \log_2\left(\frac{2}{6}\right)\right) + \left(-\frac{4}{6} * \log_2\left(\frac{4}{6}\right)\right) = 0,9183$$

$$b. \text{Entropy}(Y_a) = \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right) = 0,8113$$

A. Menghitung Nilai Gain tiap – tiap atribut.

- a. Gain (Total, Usia)
- $$= 0,8813 - \left(\left(\frac{2}{10} * 1\right) + \left(\frac{0}{10} * 0,0\right) + \left(\frac{2}{10} * 1\right) + \left(\frac{4}{10} * 0,8113\right) + \left(\frac{2}{10} * 0,0\right) + \left(\frac{0}{10} * 0,0\right)\right) = 0,1568$$
- b. Gain (Total, Lama rawatan)
- $$= 0,8813 - \left(\left(\frac{4}{10} * 0,8113\right) + \left(\frac{4}{10} * 1\right) + \left(\frac{2}{10} * 0,0\right) + \left(\frac{0}{10} * 0,0\right) + \left(\frac{0}{10} * 0,0\right) + \left(\frac{0}{10} * 0,0\right)\right) = 0,1568$$
- c. Gain (Total, Jenis kelamin)
- $$= 0,8813 - \left(\left(\frac{6}{10} * 0,6500\right) + \left(\frac{4}{10} * 1\right)\right) = 0,0913$$
- d. Gain (Total, Diabetes)
- $$= 0,8813 - \left(\left(\frac{5}{10} * 0,9710\right) + \left(\frac{5}{10} * 0,7219\right)\right) = 0,0349$$
- e. Gain (Total, Gagal ginjal)
- $$= 0,8813 - \left(\left(\frac{9}{10} * 1\right) + \left(\frac{1}{10} * 0,9710\right)\right) = 0,0548$$
- f. Gain (Total, Stroke)
- $$= 0,8813 - \left(\left(\frac{3}{10} * 0,9183\right) + \left(\frac{7}{10} * 0,5917\right)\right) = 0,1916$$
- g. Gain (Total, Hipoglikemia)
- $$= 0,8813 - \left(\left(\frac{6}{10} * 0,9183\right) + \left(\frac{4}{10} * 0,8113\right)\right) = 0,0058$$

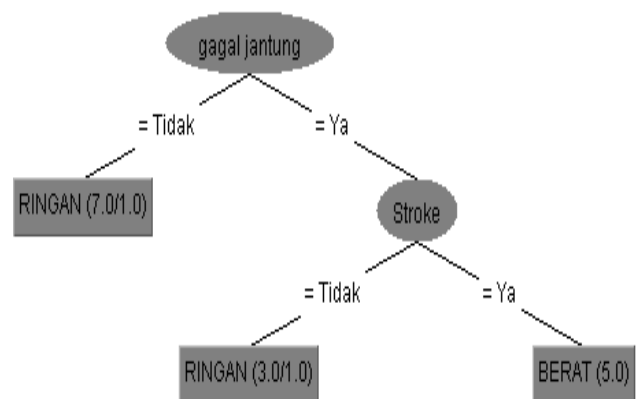
Hasil perhitungan tersebut dapat dilihat pada gambar berikut.

Node	Atribut	Jumlah Kasus(S)	RINGAN	BERAT	Entropy	Gain
1.1	TOTAL	10	3	7	0.8813	
	Usia					
	Remaja Akhir	2	1	1	1.0000	0.1568
	Dewasa awal	0	0	0	0.0000	
	Dewasa Akhir	2	1	1	1.0000	
	Lansia Awal	4	1	3	0.8113	
	Lansia Akhir	2	0	2	0.0000	
	Manula	0	0	0	0.0000	
	Lama rawatan					
	A	4	1	3	0.8113	0.1568
	B	4	2	2	1.0000	
	C	2	0	2	0.0000	
	D	0	0	0	0.0000	
	E	0	0	0	0.0000	
	F	0	0	0	0.0000	
	Jenis kelamin					
	Pria	6	1	5	0.6500	0.0913
	Wanita	4	2	2	1.0000	
	Diabetes					
	Tidak	5	2	3	0.9710	0.0349
	Ya	5	1	4	0.7219	
	Gagal Ginjal					
	Tidak	9	3	6	0.9183	0.0548
	Ya	1	0	1	0.0000	
	Stroke					
	Tidak	3	2	1	0.9183	0.1916
	Ya	7	1	6	0.5917	
	Hipoglikemia					
	Tidak	6	2	4	0.9183	0.0058
	Ya	4	1	3	0.8113	

Gambar 5. Perhitungan Node 1.1

Diketahui nilai tertinggi adalah stroke yaitu sebesar 0,1916

Berikut adalah pohon keputusan yang telah tersusun dari hasil perhitungan diatas, maka di dapatkan pohon keputusan seperti gambar berikut ini.

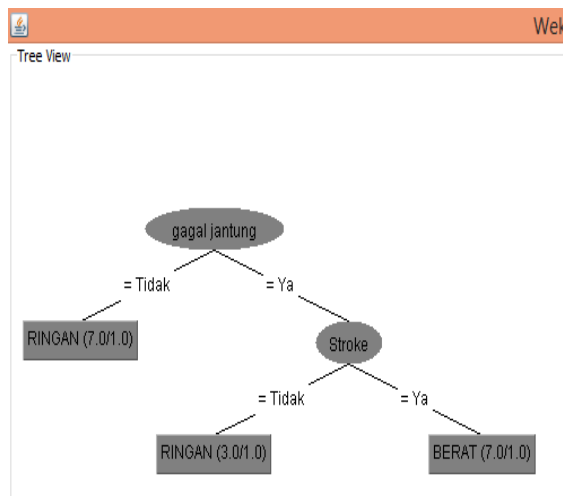


Gambar 6. Pohon keputusan akhir

Diperoleh aturan sebagai berikut:

1. Jika tidak gagal jantung maka hipertensi ringan
2. Jika gagal jantung berat maka Stroke
3. Jika tidak stroke maka hipertensi ringan
4. Jika stroke maka hipertensi Berat.

Setelah perhitungan manual selesai, maka dilakukan pengujian menggunakan software Weka. Pengujian Weka berguna untuk memastikan hasil analisa yang didapat sudah sesuai dengan yang diharapkan.



Gambar 7. Pohon Keputusan Weka

Setelah dilakukan pengujian menggunakan *software* Weka maka diperoleh aturan sebagai berikut :

1. Jika tidak gagal jantung maka hipertensi ringan
2. Jika gagal jantung berat maka Stroke
3. Jika tidak stroke maka hipertensi ringan
4. Jika stroke maka hipertensi Berat.

KESIMPULAN

Berdasarkan hasil perhitungan dan pengujian hasil, diperoleh kesimpulan sebagai berikut :

1. Dengan menggunakan metode algoritma C4.5 peneliti dapat mengklasifikasikan pola penyakit hipertensi sebagai penyakit penyerta dari penyakit gagal jantung, gagal ginjal, diabetes, stroke dan hipoglikemia.
2. Dari hasil penelitian di dapatkan informasi bahwa dalam prediksi penyakit hipertensi sebagai penyakit penyerta ,atribut yang sangat berpengaruh terhadap hipertensi adalah gagal jantung
3. Dalam penelitian ini, peneliti menggunakan *software* WEKA (*Waikato Environment for Knowledge Analysis*) sebagai *tools* atau alat yang digunakan untuk melakukan pengujian dalam rangka untuk mendapatkan pola penyakit dari hipertensi.

REFERENCES

- [1] A. A. K. Hartama, "KLASIFIKASI PENYAKIT HIPERTENSI MENGGUNAKAN ALGORITMA C4.5 STUDI KASUS RSU PROVINSI NTB," 2017.
- [2] E. Rahajeng and S. Tuminah, "Prevalensi Hipertensi dan Determinannya di Indonesia," 2007.
- [3] Profil kesehatan Kota batam, "Profil Kesehatan Tahun 2015," 2015.
- [4] A. Peña-Ayala, "Educational data mining: A survey

and a data mining-based analysis of recent works," *Expert Syst. Appl.*, vol. 41, no. 4 PART 1, pp. 1432–1462, 2014.

- [5] A. E. Pramadhani and T. Setiadi, "PENERAPAN DATA MINING UNTUK KLASIFIKASI PREDIKSI PENYAKIT ISPA (Infeksi Saluran Pernapasan Akut) DENGAN ALGORITMA DECISION TREE (ID3)," *J. Sarj. Tek. Inform. e-ISSN 2338-5197*, vol. 2, no. 1, pp. 831–839, 2014.
- [6] Y. Anisa, P. Bangun, and U. Sinulingga, "AJIAN FAKTOR PENYEBAB PENDERITA HIPERTENSI DENGAN MENGGUNAKAN ANALISIS FAKTOR DI KOTAMADYA MEDAN (Studi Kasus: RSUP H. Adam Malik)," *Saintia Mat.*, vol. 2, no. 4, pp. 333–343, 2014.
- [7] Khotimah, "Stres Sebagai Faktor Terjadinya Peningkatan Tekanan Darah Pada Penderita Hipertensi," *Eduhealth*, vol. 3, no. 2, pp. 79–83, 2013.
- [8] M. Nizar Syarif Hamidi, "Hubungan Pola Makan Dengan Kejadian Penyakit Hipertensi Di Puskesmas Kuok Tahun 2014," vol. 5, pp. 28–34, 2014.
- [9] Kementerian Pendidikan dan Kebudayaan Republik Indonesia, "Badan Pengembangan dan Pembinaan Bahasa," 2016. .
- [10] Y. Mardi, "Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219.
- [11] I. N. Syahfitri, "Penerapan Data Mining untuk Menentukan Besar Pinjaman pada Koperasi Simpan Pinjam," *Juisi*, vol. 03, no. 02, pp. 18–27, 2017.
- [12] Andri, Y. N. Kunang, and S. Murniati, "Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang," *Semin. Nas. Inform. 2013 (semnasIF 2013)*, vol. 2013, no. A, pp. 56–63, 2013.
- [13] D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4 . 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," vol. VI, no. 1, pp. 15–20, 2014.
- [14] M. H. Iswanto, A. E. Permanasari, H. A. Nugroho, J. Grafika, and N. Bulaksumur, "Pemanfaatan Teknik Data Mining Untuk Diagnosis Penyakit Tuberculosis (TBC)," *Semin. Nas. Teknol. Inf. dan Multimed. (STMIK AMIKOM)*, Yogyakarta, pp. 121–126, 2015.
- [15] S. Pujiono, A. Amborowati, M. Suyanto, and K. Kunci, "Analisis kepuasan publik menggunakan weka dalam mewujudkan," *J. DASI*, vol. 14, no. 2, pp. 45–55, 2013.
- [16] M. K. Kusriani, *Konsep dan Aplikasi Sistem Pendukung Keputusan*. Yogyakarta: Andi.